

# Weekly Report

Apr. 8<sup>th</sup>, 2018

## Done:

1. 考虑到后面计算的性能问题, 开始使用 Falcon (python)作后端; 本周打通了流程, 用 webpack 打包生成 bundle, 由 python WSG 和 falcon 来 serve 页面; 连接基于 PostgreSQL 的 PostGIS, 接收查询并返回前端;
2. 前端抽出了地图部分方便同学开发别的 deck.gl 模块; 加入火灾事件描述的视图和 hover 显示当前火灾状况的功能;
3. 上课, 一门是这边学校的研究生算法设计分析, 一门是 Coursera 上吴恩达的深度学习五门课、完成了前三周的课程及作业.

## To do:

1. 后端改成 python 后, 之前一些接口要改下代码; 此外另一同学开发的 deck.gl layer 也会将前端计算改成后端 python 加快速度, 这样可以尽快并到现有系统之中.

## 投稿总结:

今年没有顺利投稿, 原因较多:

- (a) 编程能力方面, 自己原因占多数, 马老师这边的同学中途彻底退出也是一个问题; 总得来说教训深刻, 中间问题非常多寻找答案一直在国内外同学之间求助、非常耗时; 但在此过程中也已学习了不少前后端知识与技术, 如 React 等前端全家桶, Leaflet, Deck.gl 等地图相关的前端产品, 以及 UI 框架.
- (b) 交流上, 出现了比较多的障碍; 我一开始看论文看了很久想了一些想法并和老师同学交流, 但过去将近一个月才被告知这个项目没有提供任何数据需要完全自己去找; 成员退出不久后进展缓慢, 但被指责为何不提前一个多月说缺人手; 由于出现比较多次“这么简单都不会的”指责让人已然失去继续交流的勇气. 当然现在只能靠把脸皮练厚和更多交流来解决了.  
有时候让我多找他, 要见到就抓住他聊; 但看到我的效果又说我这个没啥讲的. 这个很矛盾, 还是编程能力不过关所致.
- (c) 时间安排把握上, 编程和交流出现问题, 已大量耗时. 而没有任何可以使用的数据是另一个问题. 在之前想的一些可能的跟这个火灾数据、模型分析相关的 idea 中, 我无法找来可用的小数据集来做任何测试.
- (d) 心态方面, 由于马老师要求比较严格、而在这边村里又无法外出散心, 所以情绪经常十分消极; 但撑到现在, 也算是得到了锤炼. 这次投稿身心都受到巨大压力, 但也了解更多做人做事的方法论.

## 一些计划:

1. 学习基本知识的角度:
  - (a) 继续学习写代码;
  - (b) 跟着 Coursera 和 YouTube 视频学习深度学习基本理论以及使用的技巧, 这些课作业实践性比较强;
  - (c) 上这边的研究生课程, 主要是算法相关的
2. 下面说一写近来听到的或者看论文有关的想法. 由于个人现在努力方向(如作为博士毕业

论文到底应该往哪一块仍不明朗,因而想法会比较多一点).

首先有两点,是这边一位访学的同学回国之前问马老师“可视化的一些正处于黎明前黑暗的领域”,其中有两个是我尚且有点了解的,一个是 uncertainty, 一个是 graph drawing(另一个我完全没有了解的是体绘制领域).

- (a) uncertainty 可视化比较宽泛,因为本身对于 uncertainty 就定义的很宽;从当初对于火灾数据的调研、处理以及和这边实验室同学交流时候了解到了 uncertainty propagation 这个东西. 它不一定是在科学数据条件下的,在 VAST 数据里也会有出现;这种传播也类似是一种时间序列、evolutionary 的过程,它会存在分析的不同环节、一般也有明确的触发条件(上下文信息、各种条件的互相关联).
- (b) 在图或者网络这方面个人觉得比较有意思,图结构预测话题比较新颖但,实际生活中的需求也大;此外,在我读过的其他图类型的论文中,社交网络尽管一直很热但是最近做社交网络图可视化的也不多,多是进行了其他的转换和编码、或者是十分小的社交图可视化. 另一方面,我们之前提到的对于图比如 Entity-Relationship 之类的分析或者用于预测,记得 EuroVis 有文章以及像很久以前马老师这边的 OntoVis,但是在我所知的可视化领域做的也不多.
- (c) 机器学习与可视化结合方面,这里又分成几种
  - (1) 可视化模型过程展示、帮助分析与决策,现在 TVCG 游戏数据在审的论文就是一种类型,这比较需要实际的应用场景驱动.
  - (2) 通过多个简单函数(比如决策树、回归)逼近复杂模型来解读(如这篇 KDD16: “Why Should I Trust You?” Explaining the Predictions of Any Classifier). 这里又分成两种,一种是类似**模型的行为分析**. 两篇做 ensemble learning 可视化的文章有这种思想,多个子模型分析、调参,观察结果的优劣;一类是结合实际的**语义信息**、并且具体到这个特征、这个参数起了什么作用. 做法是多处采样、用简单模型逼近、然后组合成一个大的逼近模型的组,用这个模型组来解读本来复杂的模型. 马哥曾经说过有一个叫 rule extraction 的领域和这个非常相似,但似乎还没有做这个的可视化文章.
  - (3) 深度学习、概率图模型在实际生产应用中比较火热,近两年可视化文章也有数篇,但是由于我在技术细节方面仍在学习中、不了解具体情况所以我仍然需要继续了解以后继续;当然上面的(2)提到的第二种方法其实也可以拿来用,比如在医疗领域,为什么这写症状、这些病症导致了一种病,用上述方法可以作为一种解决方案.

### 3. 投稿计划

- (a) 手头这个项目我后面还要和同学老师细聊. 之前设想的基于图方面的推荐是比较切 funding 的一个思路,希望通过构建多属性边的 graph 来推荐相似火灾这是基于模型的;火灾数据 uncertainty 的展示、交互,小数据样本的计算、展示功能,在 uncertainty 传播和具体可视化交互上还未有详细考虑. 以上两点,下周实验室有关同学来齐了,我会细问一下他们的相关知识.  
上面提的数据准备上,最近会有新同学负责,进程会快一点.
- (b) 去年投的 VIS\CHI 论文,数据和总体算法(替换类似方法)方面需要换,系统设计这一块改动不用很大,但此文章贡献点微弱.

### 4. 其他一些工作

四月份需要完成审稿和大黑书的修订. 两者都已经初步浏览和注释,后面会尽快完善.